

# IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF

François Ehrenmann<sup>1</sup>, Quentin Kaas<sup>1</sup> and Marie-Paule Lefranc<sup>1,2,\*</sup>

<sup>1</sup>IMGT®, the international imMunoGeneTics information system®, Université Montpellier 2, Laboratoire d'ImmunoGénétique Moléculaire LIGM, Institut de Génétique Humaine IGH, UPR CNRS 1142, 141 rue de la Cardonille, 34396 Montpellier cedex 5 and <sup>2</sup>Institut Universitaire de France, 103 Bd St Michel, 75005 Paris, France

Received September 11, 2009; Revised October 9, 2009; Accepted October 12, 2009

## ABSTRACT

**IMGT/3Dstructure-DB is the three-dimensional (3D) structure database of IMGT®, the international ImMunoGenetics information system® that is acknowledged as the global reference in immunogenetics and immunoinformatics.** IMGT/3Dstructure-DB contains 3D structures of immunoglobulins (IG) or antibodies, T cell receptors (TR), major histocompatibility complex (MHC) proteins, antigen receptor/antigen complexes (IG/Ag, TR/peptide/MHC) of vertebrates; 3D structures of related proteins of the immune system (RPI) of vertebrates and invertebrates, belonging to the immunoglobulin and MHC superfamilies (IgSF and MhcSF, respectively) and found in complexes with IG, TR or MHC. IMGT/3Dstructure-DB data are annotated according to the IMGT criteria, using IMGT/DomainGapAlign, and based on the IMGT-ONTOLOGY concepts and axioms. IMGT/3Dstructure-DB provides IMGT gene and allele identification (CLASSIFICATION), region and domain delimitations (DESCRIPTION), amino acid positions according to the IMGT unique numbering (NUMEROTATION) that are used in IMGT/3Dstructure-DB cards, results of contact analysis and renumbered flat files. In its Web version, the IMGT/DomainGapAlign tool analyses amino acid sequences, per domain. Coupled to the IMGT/Collier-de-Perles tool, it provides an invaluable help for antibody engineering and humanization design based on complementarity determining region (CDR) grafting as it precisely defines the standardized framework regions (FR-IMGT) and

**CDR-IMGT. IMGT/3Dstructure-DB and IMGT/DomainGapAlign are freely available at <http://www.imgt.org>.**

## INTRODUCTION

IMGT/3Dstructure-DB (1) is the three-dimensional (3D) structure database of IMGT®, the international ImMunoGenetics information system® (<http://www.imgt.org>) (2) that is acknowledged as the global reference in immunogenetics and immunoinformatics (3). IMGT/3Dstructure-DB contains 3D structures of (i) antigen receptors that comprise immunoglobulins (IG) or antibodies and T cell receptors (TR), (ii) major histocompatibility complex (MHC) proteins of class I (MHC-I) and class II (MHC-II), (iii) peptide/MHC (pMHC) complexes (pMHC-I, pMHC-II), (iv) antigen receptor/antigen complexes (IG/Ag, TR/pMHC) and (v) related proteins of the immune system (RPI) belonging to the immunoglobulin and MHC superfamilies (IgSF and MhcSF, respectively) and found in complexes with IG, TR or MHC. Major breakthroughs characterize IMGT® and, therefore, IMGT/3Dstructure-DB: a standardized identification (IMGT keywords), a standardized nomenclature (IMGT gene and allele names), a standardized description (IMGT labels) and a standardized numerotation (IMGT unique numbering). These rules are described in the IMGT Scientific chart and are derived from the concepts of identification, classification, description and numerotation, themselves generated from the IDENTIFICATION, CLASSIFICATION, DESCRIPTION and NUMEROTATION axioms of the Formal IMGT-ONTOLOGY, the first ontology in immunogenetics and immunoinformatics (4,5).

The IMGT/3Dstructure-DB structural data are extracted from the Protein Data Bank (PDB) (6,7) and

\*To whom correspondence should be addressed. Tel: +33 4 99 61 99 65; Fax: +33 4 99 61 99 01; Email: marie-paule.lefranc@igh.cnrs.fr  
Present address:

Quentin Kaas, The University of Queensland, Institute for Molecular Bioscience, Brisbane QLD 4072, Australia.

annotated according to the IMGT-ONTOLOGY concepts of classification, using internal tools and IMGT/DomainGapAlign. Thus, IMGT/3Dstructure-DB provides the closest genes and alleles that are expressed in the amino acid sequences of the 3D structures, by aligning these sequences with the IMGT domain reference directory. This directory contains, for the antigen receptors, amino acid sequences of the domains encoded by the constant (C) genes and the translation of the germline variable (V) and joining (J) genes. The identified genes are classified based on the IMGT nomenclature of IG and TR genes and alleles (8,9) that was approved in 1999 by the Human Genome Organisation (HUGO) Nomenclature Committee (HGNC) (10), entered in IMGT/GENE-DB (11) and endorsed by the World Health Organization (WHO)-International Union of Immunological Societies (IUIS) Nomenclature Committee (12,13). Entrez Gene (14) at the National Center for Biotechnology Information (NCBI), Vertebrate Genome Annotation (Vega) (15) at the Wellcome Trust Sanger Institute and Ensembl at the European Bioinformatics Institute (EBI) currently use IMGT nomenclature.

IMGT/3Dstructure-DB provides the IMGT Colliers de Perles (16,17) that are 2D graphical representations based on the IMGT unique numbering, the key IMGT-ONTOLOGY concept of numerotation. IMGT Colliers de Perles are currently available for three domain types: V type (that includes the V-DOMAIN of IG and TR, and V-LIKE-DOMAIN of the IgSF other than IG and TR) (18), C type (that includes the C-DOMAIN of IG and TR, and C-LIKE-DOMAIN of the IgSF other than IG and TR) (19) and G type (that includes the G-DOMAIN of MHC, and G-LIKE-DOMAIN of the MhcSF other than MHC) (20).

IMGT-ONTOLOGY concepts have been crucial in bridging the gap between sequences and 3D structures in IMGT/3Dstructure-DB (21,22). Since 2008, amino acid sequences of monoclonal antibodies (mAb, suffix -mab) and of fusion proteins for immune applications (FPIA, suffix -cept) from the World Health Organization/International Nonproprietary Name (WHO/INN) have been entered in IMGT®. Being able to analyse sequences and 3D structures with the same concepts of classification (nomenclature), description (labels) and numerotation (IMGT unique numbering) facilitates comparative studies of antigen receptors and FPIA, and of their constitutive chains, even if 3D structures are not yet available. Results from IMGT/DomainGapAlign and the associated IMGT/Collier-de-Perles tool are widely used by researchers, particularly for antibody engineering and humanization design.

## IMGT/3DStructure-DB DATABASE

Besides an increase in the number of IMGT/3Dstructure-DB coordinate files (634 in 2004, 1649 in 2009), the database itself was enhanced with (i) new data types (IgSF other than IG and TR, MhcSF other than MHC), (ii) novel functionalities that include chain sequences in

IMGT format, contact analysis of paratope/epitope, 'IMGT pMHC contact sites' (23,34) and 'IMGT numbering comparison' and (iii) improvement of the IMGT/DomainGapAlign tool that allows the identification of the closest gene(s) and allele(s) expressed in IMGT/3Dstructure-DB protein sequences and IG and TR variable domains, by comparison with the IMGT domain directory, and the display of amino acid differences.

IMGT/3Dstructure-DB is queried through a user-friendly interface and different displays for the results can be chosen from the database home page, as previously described (1). Each entry in the database is detailed in an IMGT/3Dstructure-DB card. Eight tabs are available at the top of each card: 'Chain details', 'Contact analysis', 'Paratope and epitope', '3D visualization Jmol or QuickPDB', 'Renumbered IMGT file', 'IMGT numbering comparison', 'References and links' and 'Printable card'. First, we will describe 'Chain details', and then a selection of three features related to 'Contact analysis' (IMGT Collier de Perles on two layers with hydrogen bonds, IMGT pMHC contact sites, IMGT/3Dstructure-DB Domain pair contacts) and the IMGT/DomainGapAlign results in its Web version (Figure 1).

## Statistics

In September 2009, the IMGT/3Dstructure-DB database manages 1649 coordinate files that correspond to 1773 receptors (Table 1).

'Other species' refers to receptors from species other than *Homo sapiens* (human) and *Mus musculus* (mouse), and also to engineered proteins (humanized, chimeric or synthetic). Thus, 'Other species' refers, for IG, to *Camelus dromedarius* (36), *Lama glama* (10), *Rattus norvegicus* (14), *Ginglymostoma cirratum* (6), *Orectolobus maculatus* (5), *Cricetinae* gen. sp. (1), *Oryctolagus cuniculus* (1), *Cavia porcellus* (1), humanized (23), chimeric (29) and synthetic (18); for TR, to *R. rattus* (1); for MHC, to *R. norvegicus* (3); and for RPI, to *R. norvegicus* (9), *Hyalophora cecropia* (1), *Branchiostoma floridae* (2) and chimeric (2). A total of 318 different genes (corresponding to 415 alleles) were identified in V-DOMAIN: 266 for IG (155 IGHV, 93 IGKV and 18 IGLV) and 52 for TR (29 TRAV, 18 TRBV, 3 TRDV and 2 TRGV). IMGT/3Dstructure-DB contains 920 antigen receptor/antigen complexes that comprise 588 IG/Ag (141 human, 393 mouse and 54 'other species') and 38 TR/pMHC (22 human, 16 mouse). There are 294 pMHC complexes (192 human, 99 mouse and 3 other species) (Table 1).

## IMGT/3Dstructure-DB 'Chain details'

The 'Chain details' (Figure 1A) comprises information, first on the chain itself, and then on each domain, starting from the N-terminal end.

IMGT chain annotation includes:

- Chain ID;
- Chain length;
- IMGT chain description, with IMGT labels indicating the constitutive domains and their delimitations [for

Chain ID	1oga_E
Chain length	252
IMGT chain description	= V-BETA-1(1-110) + C-BETA-1(111-252)
Chain sequence	<pre>V-REGION MVDGIGITQSPKYLFRKEQGNVTLSCENQLNHADAMYRQDPGQGLRLIYYSQIVNDQKGDIAEYGVSRSKEKSFPLTVTSQAKNPNTAF IN-A[ J-REGION ]I YLQASSRSRSEQYFGPQTPLTVTEULKVNWPPEAVFEPSEAEISHQKATLVCLATGFYDPIVELSMWVNGKEVHSVGSTDPQPLKEQ ON PALNDNSRYLSLSSRLRVSATRWQNPFRHRCQVQFYGLSENDENTQDRAKPVQTQIVSAEAWGRADQDRGGGCD</pre>
	<i>Sequence in FASTA format</i> <i>Sequence in IMGT format</i>
IMGT domain description	V-BETA
IMGT gene and allele name	TRBV1 9*01 (100.00%)(Human) <a href="#">Alignment details</a>
IMGT gene and allele name	TRBJ2-7*01 (100.00%)(Human) <a href="#">Alignment details</a>
2D representation	<a href="#">IMGT Collier de Perles</a> or <a href="#">IMGT Collier de Perles on 2 layers</a>
Contact analysis	<a href="#">Domain contacts (overview)</a>
CDR-IMGT lengths	[5.6.11]
Sheet composition	[A' B D E] [A' C C' F G]
	<pre>[ CDR1 ] DGGITQSPKYLFRKEQGNVTLSCENQLNH.....DAWYWRQDPGQGLRLIYYSQI....VNFQKGDIA.EGYSVSR.EKKESPLT [ CDR3 ] VTSQAKNPTAFYLCAASSRSR...SEQYFGPQTPLTV</pre>
	<a href="#">IMGT/DomainGapAlign results</a>
IMGT domain description	C-BETA-1
IMGT gene and allele name	TRBC2*01 (99.20%)(Human), TRBC2*02 (99.20%)(Human) <a href="#">Alignment details</a>
2D representation	<a href="#">IMGT Collier de Perles</a> or <a href="#">IMGT Collier de Perles on 2 layers</a>
Contact analysis	<a href="#">Domain contacts (overview)</a>
Sheet composition	[A B D E] [C F G]
	.EDLKWNWPPEAVFEPSEAEISH..TQAKATLVCLATGFYP..DHVELSMWVNGKEVHS..GVSTDQPLKEQPAL.NDSRYLSLSSRLRV SATWQ.NPRNHRFCQVQFYGLSENDENTQDRAKPVQTQIVSAEAWGRADQDRGGGCD
	<a href="#">IMGT/DomainGapAlign results</a>

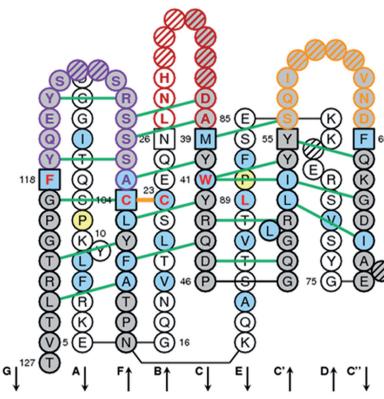
**A**  
**THANK YOU**  
**for using IMGT/3Dstructure-DB**

THE  
INTERNATIONAL  
IMMUNOGENETICS  
INFORMATION SYSTEM®



<http://www.imgt.org>

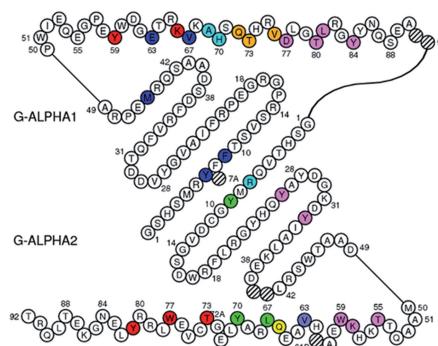
IMGT Collier de Perles : Homo sapiens (Human) TRBV V-BETA DOMAIN [D1] from JM22 (1oga\_E)  
CDR-IMGT lengths [5.6.11]



### MHC chain: 1bd2\_A

Peptide	Contact sites	G-ALPHA1	G-ALPHA2
1 L	C1	59 66	73 77 81
2 L	C3	7 9 45 63 67	
3 F	C4		9 67 70
4 G	C5		
5 Y	C6		66
6 P	C8	69 70	7
7 V	C9		63
8 Y	C10	72 73 76	
9 V	C11	77 80 81 84	26 33 55 58 59

### IMGT Collier de Perles with pMHC contact sites



### C

IMGT/3Dstructure-DB Domain pair contacts  
Contacts of **V-BETA 1oga\_E** with **Domain Chain** **1oga\_C** (Ligand)

Atom contact types Atom contact categories  
 Non covalent  Covalent  
 Polar  Disulfide  
 Hydrogen bond  (RB) Backbone/backbone  
 Non polar  (SS) Side chain/side chain  
  (BS) Backbone/side chain  
  (SB) Side chain/backbone  
 Check all  Check all  
 Uncheck all  Uncheck all

#### Summary:

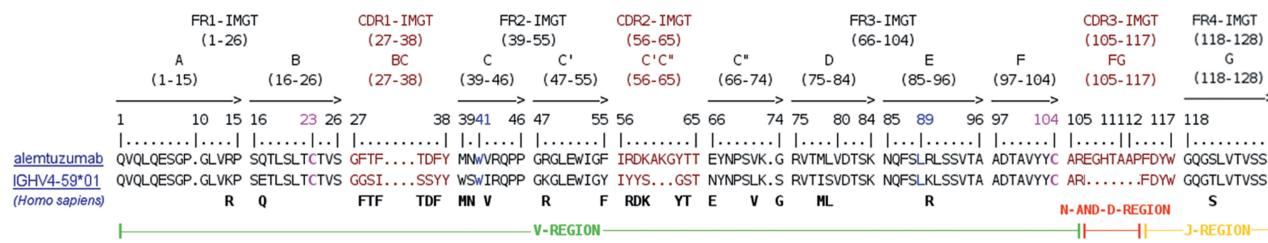
Residue contacts	Number of residues	Atom contact types
Total From 1 From 2	63 16 3	Polar Hydrogen Non Polar

#### List of the Residue@Position pair contacts:

Click TgB® for IMGT Residue@Position cards

Order	IMGT Num	Residue	Domain	Chain	Order	IMGT Num	Residue	Domain	Chain	Total	Polar	Hydrogen	Non Polar
RQP	37	ASP	D	V-BETA 1oga_E	RQP	8	THR	T	1oga_C	12	3	1	9
RQP	57	GLN	Q	V-BETA 1oga_E	RQP	4	GLY	G	1oga_C	6	2	1	4
RQP	57	GLN	Q	V-BETA 1oga_E	RQP	5	PHE	F	1oga_C	12	3	0	9
RQP	57	GLN	Q	V-BETA 1oga_E	RQP	6	VAL	V	1oga_C	20	4	1	16
RQP	58	ILE	I	V-BETA 1oga_E	RQP	6	VAL	V	1oga_C	2	0	0	2
RQP	58	ILE	I	V-BETA 1oga_E	RQP	8	THR	T	1oga_C	3	0	0	3
RQP	107	ARG	R	V-BETA 1oga_E	RQP	5	PHE	F	1oga_C	3	0	0	3
RQP	108	ARG	R	V-BETA 1oga_E	RQP	6	VAL	V	1oga_C	1	0	0	1
RQP	109	ARG	R	V-BETA 1oga_E	RQP	7	PHE	F	1oga_C	6	0	0	6
RQP	110	SER	S	V-BETA 1oga_E	RQP	5	PHE	F	1oga_C	4	0	0	4
RQP	110	SER	S	V-BETA 1oga_E	RQP	6	VAL	V	1oga_C	9	4	0	5
RQP	113	SER	S	V-BETA 1oga_E	RQP	5	PHE	F	1oga_C	5	0	0	5

### Alignment with the closest genes and alleles from the IMGT domain directory



### E

**Figure 1.** (A) IMGT/3Dstructure-DB ‘Chain details’. (B) IMGT Collier de Perles on two layers with hydrogen bonds. (C) IMGT pMHC contact sites of human HLA-A\*0201 MHC-I and a 9-amino acid peptide (1bd2). (D) IMGT/3Dstructure-DB Domain pair contacts between a V-BETA domain and a ligand (1oga). (E) IMGT/DomainGapAlign Web results of alemtuzumab.

- example, in Figure 1A: TR-BETA-1 = V-BETA (1–110) + C-BETA (111–239)]; and
- Chain sequence, with delimited regions (for example, V-REGION, N-AND-REGION and J-REGION for a V-BETA domain) and links to *Sequence in FASTA format* and *Sequence in IMGT format*.

IMGT domain annotation includes:

- IMGT domain description;
- IMGT gene and allele name, with percentage of identity and species, and a link to *Alignment details* (V and J genes and alleles are given, in the case of a V domain, as in Figure 1A);
- 2D representation, with links to *IMGT Collier de Perles* (on one layer) and, for V and C type domains, to *IMGT Collier de Perles on two layers*;
- Contact analysis, with a link to *Domain contacts (overview)*;

**Table 1.** Number of receptors and complexes in IMGT/3Dstructure-DB

	Number per species			
	<i>Homo sapiens</i>	<i>Mus musculus</i>	Other species	Total
<b>Receptor type</b>				
IG	293	655	144	1092
TR	48	40	1	89
MHC	251	133	3	387
RPI	168	23	14	205
Total	760	851	162	1773
<b>Complex type<sup>a</sup></b>				
Antigen receptor/antigen				
IG/Ag	141	393	54	588
TR/pMHC				
TR/pMHC-I	17	10	0	27
TR/pMHC-II	5	6	0	11
peptide/MHC				
pMHC-I	155	90	3	248
pMHC-II	37	9	0	46
Total	355	508	57	920

<sup>a</sup>Complexes involving RPI are not shown.

**Table 2.** Number of IMGT/3Dstructure-DB data for IMGT Collier de Perles, Contact analysis and IMGT/DomainGapAlign

	Number <sup>a</sup>	IG, TR, IgSF RPI <sup>b</sup>	MHC, MhcSF RPI <sup>c</sup>
IMGT Collier de Perles			
On one layer	9000	V, C	G, C
On two layers with hydrogen bonds	22 000	V, C	nr
Contact analysis			
IMGT pMHC contact sites	15 000	nr	Peptide – G
IMGT/3Dstructure-DB Domain pair contacts	45 000	(V or C) – (V or C) (V or C) – (Ligand)	(G or C) – (G or C) (G or C) – (Ligand)
IMGT Residue@Position cards	1500 000	V, C	G, C
IMGT/DomainGapAlign			
IMGT/DomainGapAlign results	9000	V, C	G, C

<sup>a</sup>Numbers (September 2009) have been rounded off.

<sup>b</sup>IgSF RPI: IgSF proteins other than IG and TR.

<sup>c</sup>MhcSF RPI: MhcSF proteins other than MHC. MHC and MhcSF RPI also belong to IgSF RPI, by their C-like domain, if present.

V, V type; C, C type; G, G type; nr, not relevant.

- CDR-IMGT lengths: lengths of the three CDR, CDR1-IMGT, CDR2-IMGT and CDR3-IMGT, shown between brackets and separated by dots, for example [5.6.11], this information being determined by IMGT/DomainGapAlign;
- Sheet composition: strands of each layer being indicated between brackets, for example [A'BDE] [A"CC'C"FG], this information being deduced from the PDB file; and
- Domain amino acid sequence with gaps according to the IMGT numbering and region delimitations (for example, CDR1, CDR2, CDR3 for a V-DOMAIN, as in Figure 1A) and a link to *IMGT/DomainGapAlign results*.

### IMGT Collier de Perles on two layers with hydrogen bonds

IMGT Colliers de Perles on two layers, available for the V and C type domains (18,19), are displayed with hydrogen bonds (green lines in Figure 1B), this information being determined from structural data extracted from the PDB files. The V type domain is made of nine antiparallel beta strands (A, B, C, C', C'', D, E, F and G) linked by beta turns (AB, CC', C''D, DE and EF) or loops (BC, C'C'' and FG), forming a sandwich of two sheets (18). The sheets are closely packed against each other through hydrophobic interactions giving a hydrophobic core and joined together by a disulfide bridge between the B-STRAND in the first sheet and the F-STRAND in the second sheet. In the IMGT unique numbering, the conserved amino acids always have the same position, for instance cysteine 23 (1st-CYS), tryptophan 41 (CONSERVED-TRP), conserved hydrophobic amino acid 89 and cysteine 104 (2nd-CYS) (18). The C type domain has similar topology and 3D structure than the V type domain, but it differs by the number of strands: it is made of seven beta strands linked by beta turns or loops, and arranged so that four strands form one sheet and three strands form a second sheet (19). In September 2009, there are about 22 000 IMGT Colliers de Perles on two layers with hydrogen bonds that correspond to 7300 V and C domains (Table 2).

### IMGT pMHC contact sites

'IMGT pMHC contact sites' graphically represent, in the IMGT Colliers de Perles for G-DOMAIN (20), the MHC amino acid positions that contact the peptide side chains in pMHC complexes (Figure 1C), and thus allow comparison of pMHC interactions (23,24). Eleven 'IMGT pMHC contact sites' were defined (C1–C11). They take into account the length of the peptides and are considered independently of the MHC class and sequence polymorphisms. All direct contacts (defined with a cutoff equal to the sum of the atom van der Waals radii and of the diameter of a water molecule) and water-mediated hydrogen bonds are taken into account for the determination of the peptide/MHC interface. Amino acid positions involved in the binding interface are filtered and then classified in the IMGT pMHC contact sites that combine contact analysis (as detailed in next paragraph) with an interaction scoring function. The score assigned to each contact is a constant value, independent on the distance between atoms [40 for direct hydrogen bond, 20 for water-mediated hydrogen bond, 20 for polar interaction and 1 for nonpolar interaction (23)], which roughly complies with the true mean energy ratio (25). The IMGT pMHC contact sites allow to compare, for the first time in a database, the pMHC contacts whatever the MHC class (I or II), whatever the G domain (G-ALPHA1, G-ALPHA2, G-ALPHA or G-BETA) and whatever the peptide length. This structural information is an important asset in peptide vaccine design, particularly for peptide binding to MHC-II as it allows to clearly identify and to visualize amino acids in the groove.

### IMGT/3Dstructure-DB Domain pair contacts

'IMGT/3Dstructure-DB Domain pair contacts' (Figure 1D) measure contacts between structural units (domains or ligand). Contacts are obtained in IMGT/3Dstructure-DB by a local program written in C in which atoms are considered to be in contact when no water molecule can take place between them. The cutoff using the diameter of a water molecule allows to identify the van der Waals interactions. For the identification of hydrogen bonds, more stringent criteria are considered: the distance between the donor and acceptor atoms must be below or equal to 3.2 Å; the angle C–A–D, defined by the carbon connected to the acceptor atom (C), the acceptor atom (A) and the donor atom (D), must be above 90°; the angle D–H–A, defined by the donor atom (D), the hydrogen atom given by the donor atom (H) and the acceptor atom (A), must be above 140°. Two cases arise: (i) if the hydrogen coordinates can be directly computed (hydrogen linked to the backbone nitrogen, to the nitrogen of the tryptophan side chain, etc.), the angle is directly measured; (ii) in the case of the donor being linked to a single heavy atom, the position of the hydrogen cannot be directly computed (because of the free rotation around the axis defined by the donor and this heavy atom). In that case the hydrogen atom is supposed to be coplanar with the carbon linked to the donor atom, the donor (D) and the acceptor (A). This allows us to compute the D–H–A angle by a simple

geometric equation. The coplanar assumption is acceptable given the free rotation mentioned above. The angles' and distances' values used for the computation are derived from (25). The atom contact types are identified as 'noncovalent', 'polar', 'hydrogen bond', 'nonpolar', 'covalent' or 'disulfide' and the atom contact categories as '(BB) Backbone/backbone', '(SS) Side chain/side chain', '(BS) Backbone/side chain' and '(SB) Side chain/backbone'. As an example, Figure 1D shows the 'IMGT/3Dstructure-DB Domain pair contacts' between a V-BETA domain and its ligand (a peptide bound to a MHC). Twelve domain pair contacts, involving 11 different residues, 6 from V-BETA and 5 from the peptide, are listed with information on the atom contact types. Positions that belong to the CDR-IMGT are highlighted as they contribute to the specificity of the antigen receptors: red (CDR1-IMGT), orange (CDR2-IMGT) and purple (CDR3-IMGT), according to the IMGT color menu (same colors as in Figure 1B). In September 2009, IMGT/3Dstructure-DB contains ~45 000 domain pair contacts (Table 2).

The atom contact types and categories for each amino acid are provided in 'IMGT Residue@Position cards'. A Residue@Position is defined by the position numbering (for example 23), according to the IMGT unique numbering, the residue name in the three-letter abbreviation and in the one-letter abbreviation for amino acids [e.g. CYS (C)], the IMGT domain description (e.g. V-BETA) and the IMGT chain ID (e.g. 1loga\_E). 'IMGT Residue@Position card' can be obtained by clicking on 'R@P' in 'IMGT/3Dstructure-DB Domain pair contacts' (Figure 1D), or on any residue in the 'IMGT Collier de Perles on one layer' or in the amino acid sequence. In September 2009, IMGT/3Dstructure-DB contains about 1500 000 IMGT Residue@Position contacts (Table 2).

### IMGT/DomainGapAlign

IMGT/DomainGapAlign, used internally for the analysis of amino acid sequences in IMGT/3Dstructure-DB, is also available on the Web. The tool aligns the user amino acid sequences with the IMGT domain reference directory, identifies the closest V, C and G domains, creates gaps according to the IMGT unique numbering and highlights differences with the closest reference(s). For an antibody V domain sequence, IMGT/DomainGapAlign identifies the closest germline V-REGION and J-REGION, and provides a delimitation of the strands, framework regions (FR-IMGT) and CDR-IMGT (Figure 1E). The IMGT gene and allele name of the closest sequence(s) from the IMGT domain reference directory is provided with a percentage of identity and a Smith–Waterman score. Regions and domains are highlighted using the IMGT Color menu. Several sequences can be analysed simultaneously and users can choose how many alignments to display for each sequence. The IMGT Colliers de Perles are generated from the sequences provided by IMGT/DomainGapAlign tool.

## CONCLUSION

The IMGT standardization has allowed to build a unique frame for the comparison of the IG, TR and MHC structural data and interactions with their ligands. That standardization has been extended to the IgSF and MhcSF proteins that involve V, C and/or G type domains (26–29) and used, as an example, to develop a binding prediction method for MhcSF G type domain and beta2-microglobulin (30). IG/Ag and TR/pMHC complexes can be queried from the IMGT/3Dstructure-DB home page. IMGT/3Dstructure-DB data are particularly useful in antibody engineering and humanization design. Indeed they allow to precisely define and to easily compare amino acid sequences of the FR-IMGT and CDR-IMGT, between the nonhuman (mouse, rat, etc.) and the closest human V domains. A recent analysis performed on humanized antibodies used in oncology underlines the importance of a correct delimitation of the CDR to be grafted (31). IMGT3Dstructure-DB facilitates the identification of potential immunogenic residues at given positions in chimeric or humanized antibodies, including those of the constant domains (32,33). Moreover, the low potential immunogenicity of nonhuman primate antibodies and their potential use as therapeutics in humans has been shown by this approach (34,35). This is particularly important for nonhuman primate antibodies neutralizing *Bacillus anthracis* and the anthrax lethal toxin or the ricin toxin that could be used as therapeutics in ricin intoxication and against bioweapons (36,37). These therapeutic applications emphasize the importance of the IMGT/3Dstructure-DB standardized approach that bridges the gap between sequences and 3D structures whatever the species.

## AVAILABILITY AND CITATION

Authors who use the IMGT/3Dstructure-DB database and IMGT/DomainGapAlign are encouraged to cite this article and to quote the IMGT® Home page URL, <http://www.imgt.org>.

## ACKNOWLEDGEMENTS

We thank Véronique Giudicelli, Patrice Duroux and Gérard Lefranc for helpful discussion. We are grateful to all the IMGT® users from academic and industrial laboratories who help promoting standardization.

## FUNDING

IMGT® is funded by the Centre National de la Recherche Scientifique (CNRS), the Ministère de l'Enseignement Supérieur et de la Recherche (MESR) (Université Montpellier 2 Plan Pluri-Formation, Institut Universitaire de France), Agence Nationale de la Recherche (ANR-06-BIOSYS-005-01, FLAVORES), Région Languedoc-Roussillon (G PTR, GEPETOS), GIS IBiSA Infrastructures en Biologie Santé et Agronomie (2008-37). Funding for open access charge: CNRS.

*Conflict of interest statement.* None declared.

## REFERENCES

- Kaas,Q., Ruiz,M. and Lefranc,M.-P. (2004) IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. *Nucleic Acids Res.*, **32**, D208–D210.
- Lefranc,M.-P., Giudicelli,V., Ginestoux,C., Jabado-Michaloud,J., Folch,G., Bellahcene,F., Wu,Y., Gemrot,E., Brochet,X., Lane,J. *et al.* (2009) IMGT®, the international ImMunoGeneTics information system®. *Nucleic Acids Res.*, **37**, D1006–D1012.
- Warr,G.W., Clem,L.W. and Söderhall,K. (2003) Editorial—the international ImMunoGeneTics database IMGT. *Dev. Comp. Immunol.*, **27**, 1–2.
- Giudicelli,V. and Lefranc,M.-P. (1999) Ontology for immunogenetics: IMGT-ONTOLOGY. *Bioinformatics*, **15**, 1047–1054.
- Duroux,P., Kaas,Q., Brochet,X., Lane,J., Ginestoux,C., Lefranc,M.-P. and Giudicelli,V. (2008) IMGT-Kaleidoscope, the Formal IMGT-ONTOLOGY paradigm. *Biochimie*, **90**, 570–583.
- Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Bourne,P.E., Addess,K.J., Bluhm,W.F., Chen,L., Deshpande,N., Feng,Z., Fleri,W., Green,R., Merino-Ott,J.C., Townsend-Merino,W. *et al.* (2004) The distribution and query systems of the RCSB Protein Data Bank. *Nucleic Acids Res.*, **32**, D223–D225.
- Lefranc,M.-P. and Lefranc,G. (2001) *The Immunoglobulin FactsBook*. Academic Press, London, UK, pp. 1–458.
- Lefranc,M.-P. and Lefranc,G. (2001) *The T Cell Receptor FactsBook*. Academic Press, London, UK, pp. 1–398.
- Wain,H.M., Bruford,E.A., Lovering,R.C., Lush,M.J., Wright,M.W. and Povey,S. (2002) Guidelines for human gene nomenclature. *Genomics*, **79**, 464–470.
- Giudicelli,V., Chaume,D. and Lefranc,M.-P. (2005) IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.*, **33**, D256–D261.
- Lefranc,M.-P. (2007) WHO-IUIS Nomenclature Subcommittee for immunoglobulins and T cell receptors report. *Immunogenetics*, **59**, 899–902.
- Lefranc,M.-P. (2008) WHO-IUIS Nomenclature Subcommittee for immunoglobulins and T cell receptors report August 2007. 13th International Congress of Immunology, Rio de Janeiro, Brazil. *Dev. Comp. Immunol.*, **32**, 461–463.
- Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **35**, D26–D31.
- Loveland,J. (2005) VEGA, the genome browser with a difference. *Brief Bioinform.*, **6**, 189–193.
- Ruiz,M. and Lefranc,M.-P. (2002) IMGT gene identification and Colliers de Perles of human immunoglobulins with known 3D structures. *Immunogenetics*, **53**, 857–883.
- Kaas,Q. and Lefranc,M.-P. (2007) IMGT Colliers de Perles: standardized sequence-structure representations of the IgSF and MhcSF superfamily domains. *Curr Bioinformatics*, **2**, 21–30.
- Lefranc,M.-P., Pommié,C., Ruiz,M., Giudicelli,V., Foulquier,E., Truong,L., Thouvenin-Contet,V. and Lefranc,G. (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.*, **27**, 55–77.
- Lefranc,M.-P., Pommié,C., Kaas,Q., Duprat,E., Bosc,N., Guiraudou,D., Jean,C., Ruiz,M., Da Piedade,I., Rouard,M. *et al.* (2005) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev. Comp. Immunol.*, **29**, 185–203.
- Lefranc,M.-P., Duprat,E., Kaas,Q., Tranne,M., Thiriot,A. and Lefranc,G. (2005) IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily (MhcSF) G-LIKE-DOMAIN. *Dev. Comp. Immunol.*, **29**, 917–938.

21. Kaas,Q., Ehrenmann,F. and Lefranc,M.-P. (2007) IG, TR and IgSF, MHC and MhcSF: what do we learn from the IMGT Colliers de Perles? *Brief. Funct. Genomic Proteomic*, **6**, 253–264.
22. Lefranc,M.-P., Giudicelli,V., Regnier,L. and Duroux,P. (2008) IMGT, a system and an ontology that bridge biological and computational spheres in bioinformatics. *Brief Bioinformatics*, **9**, 263–275.
23. Kaas,Q. and Lefranc,M.-P. (2005) T cell receptor/peptide/MHC molecular characterization and standardized pMHC contact sites in IMGT/3Dstructure-DB. *In Silico Biol.*, **5**, 505–528.
24. Kaas,Q., Duprat,E., Tourneur,G. and Lefranc,M.-P. (2008) IMGT standardization for molecular characterization of the T cell receptor/peptide/MHC complexes. In Schoenbach,C., Ranganathan,S. and Brusic,V. (eds), *Immunoinformatics*. Series Immunomics Reviews, Vol. 1. Springer, New York, USA. Chapter 2, pp. 19–49.
25. Creighton,T.E. (1992) *Proteins: Structures and Molecular Properties*, 2nd edn. Freeman, New York, USA. ISBN 0-7617-7030-X.
26. Garapati,V.P. and Lefranc,M.-P. (2007) IMGT Colliers de Perles and IgSF domain standardization for T cell costimulatory activatory (CD28, ICOS) and inhibitory (CTLA4, PDCD1 and BTLA) receptors. *Dev. Comp. Immunol.*, **31**, 1050–1072.
27. Hansen,J.D., Pasquier,L.D., Lefranc,M.-P., Lopez,V., Benmansour,A. and Boudinot,P. (2009) The B7 family of immunoregulatory receptors: a comparative and evolutionary perspective. *Mol. Immunol.*, **46**, 457–472.
28. Bernard,D., Hansen,J.D., Du Pasquier,L., Lefranc,M.-P., Benmansour,A. and Boudinot,P. (2007) Costimulatory receptors in jawed vertebrates: conserved CD28, odd CTLA4 and multiple BTLAs. *Dev. Comp. Immunol.*, **31**, 255–271.
29. Frigou,A. and Lefranc,M.-P. (2005) MICA: standardized IMGT allele nomenclature, polymorphisms and diseases. In Pandalai,S.G. (ed.), *Recent Research Developments in Human Genetics*, Vol. 3. Research Signpost, Trivandrum, Kerala, India, pp. 95–145.
30. Duprat,E., Lefranc,M.-P. and Gascuel,O. (2006) A simple method to predict protein-binding from aligned sequences—application to MHC superfamily and beta2-microglobulin. *Bioinformatics*, **22**, 453–459.
31. Magdelaine-Beuzelin,C., Kaas,Q., Wehbi,V., Ohresser,M., Jefferis,R., Lefranc,M.-P. and Watier,H. (2007) Structure-function relationships of the variable domains of monoclonal antibodies approved for cancer treatment. *Crit. Rev. Oncol. Haematol.*, **64**, 210–225.
32. Jefferis,R. and Lefranc,M.-P. (2009) Human immunoglobulin allotypes: possible implications for immunogenicity. *mAbs*, **1**, 4, 1–7.
33. Magdelaine-Beuzelin,C., Vermeire,S., Goodall,M., Baert,F., Noman,M., Assche,G.V., Ohresser,M., Degenne,D., Dugoujon,J.M., Jefferis,R. et al. (2009) IgG1 heavy chain-coding gene polymorphism (G1m allotypes) and development of antibodies-to-infliximab. *Pharmacogenet. Genomics*, **19**, 383–387.
34. Laffly,E., Danjou,L., Condemeine,F., Vidal,D., Drouet,E., Lefranc,M.-P., Bottex,C. and Thullier,P. (2005) Selection of a macaque Fab with human-like framework regions, high affinity, and that neutralizes the protective antigen (PA) of *Bacillus anthracis*. *Antimicrob. Agents Chemother.*, **49**, 3414–3420.
35. Pelat,T., Hust,M., Laffly,E., Condemeine,F., Bottex,C., Vidal,D., Lefranc,M.-P., Dübel,S. and Thullier,P. (2007) A high affinity, human-like antibody fragment (scFv) neutralising the lethal factor (LF) of *Bacillus anthracis* by inhibiting PA-LF complex formation. *Antimicrob. Agents Chemother.*, **51**, 2758–2764.
36. Pelat,T., Bedouelle,H., Rees,A.R., Crennell,S.J., Lefranc,M.-P. and Thullier,P. (2008) Germline humanization of a non-human primate antibody that neutralizes the anthrax toxin, by in vitro and in silico engineering. *J. Mol. Biol.*, **384**, 1400–1407.
37. Pelat,T., Hust,M., Hale,M., Lefranc,M.-P., Dübel,S. and Thullier,P. (2009) Isolation of a human-like antibody fragment (scFv) that neutralizes ricin biological activity. *BMC Biotechnol.*, **9**, 60.